Project Title: Information Technology Research (ITR): Building the Tree of Life -- A National Resource for Phyloinformatics and Computational Phylogenetics

Technical/Programmatic Terms and Conditions

1. The purpose of the Information Technology Research (ITR) Program is to encourage and stimulate innovative, high risk/high return, multi-disciplinary research on the fundamental challenges facing the continued expansion and utilization of IT across the sciences and engineering, the social, economic, educational and workforce implications of IT, and the use of IT to enhance security and reduce the vulnerabilities of our society to catastrophic events, whether natural or man-made. The ITR Program is interested in fostering visionary work that could lead in the future to major advances, new and unanticipated technologies, revolutionary applications, or new ways to perform important activities. The program's main goals are to augment the nation's IT knowledge base and strengthen the IT workforce.

2. This project focuses on the design, development, and deployment of advanced information technology for phylogenetic reconstruction and analysis.

3. The Awardee will investigate major thrusts, each with ancillary lines of investigation, including:
a. Scalability: (i) Application Scalability which will focus on scaling approaches to datasets of the size of the Tree of Life—millions of elements—through novel algorithmic ideas, algorithmic engineering, and the use of high-performance computing platforms; (ii) Platform Scalability which will include local compilation of sources to ensure that the software can take best advantage of available computing resources; and (iii) Software Scalability which will demonstrate the enormous advantages of a well planned, open-source, software development effort by providing a framework in which any researcher anywhere is welcome to integrate new solution modules.
b. Evaluation: Emphasis will focus on how to develop benchmark datasets (which, for the larger datasets, will be based on realistic simulations of evolution) and evaluation procedures that can accurately predict how each computational approach will scale (in terms of accuracy and running time) when applied to datasets of millions of elements.
c. Integration and Workflows: Emphasis will focus on the integration of tools into an overall architecture that facilitates both typical workflows (such as routine phylogenetic analyses conducted in pharmaceutical laboratories) as well as discovery processes.
d. Community Infrastructure: Create an enduring IT resource available to the community and provide a unified source for all of the algorithms, tools, data, and results created under this cooperative agreement.

4. This Project will provide agreed upon deliverables, including:
a. The platform: machines and database.
b. The software suite: a framework for integration of additional modules, a population of reconstruction and analysis modules, a high-level graphical environment supporting workflows, simulation software (generating DNA sequences under simulated evolution), and evaluation tools, both for evaluating the performance of algorithms and for evaluating the quality of analyses; all open-source and packaged to download and install easily under the three main OS (Linux, Mac, and Windows).
c. Research papers: in algorithm design and analysis, algorithm engineering, database design and bioinformatics, software architecture, computational biology, molecular evolution, and phylogenetics.
d. Educational components: a comprehensive web site for research, education, and public outreach; educational modules for high school; a museum exhibit featuring evolution and phylogenetic reconstruction. The educational modules will be field-tested during summer programs with volunteer high-school students and teachers.

5. The Awardee agrees to adhere to the Project Goals and Objectives and the schedule of deliverables as delineated in the proposal, as revised in writing through mutual discussions and agreements between the Principals Investigators, Awardee, and NSF. The agreed upon Project Goals and Objectives may be accessed at http://www.nsf.gov/home/crssprgm/itr/pdf/itr_tree_life2.pdf .

6. The Awardee agrees to submit the required NSF Annual Progress Report and to include in that report
a. A discussion of the progress made in achieving the annual Project Goals and Objectives for the year under review. The discussion should address any emerging technical, managerial and financial issues, as well as proposed steps to resolve such issues. In addition, the report should include any recommended changes in the goals and objectives for the next year with justification for the recommendations.
b. A discussion and statistical information regarding performance of the scope of work achievements as related to actual fund expenditures, annual and cumulative, in the level of detail for which the Awardee has projected the succeeding year's costs, and a summary of the project budget in NSF format. .

7. The Awardee agrees to participate in a formal project review toward the end of the second year of this award that will contribute to the evaluation of project performance and management, including an assessment of progress made in achieving proposed goals and objectives and will assist in the determination of future directions for the project. The Awardee understands that additional project reviews may be conducted in later years of the project as needed to facilitate NSF's oversight role. To review the ITR site review protocol, please go to the ITR website at http://www.nsf.gov/home/crssprgm/itr/pdf/itr_site_visit.pdf.

8. The Awardee agrees to develop and maintain a web page that describes the project goals and objectives and highlights major accomplishments and related events including such items as the integration of research and education, publications, software, press, meetings and other useful links. .

9. The Awardee understands it is very important that the Awardee and NSF work together to publicize the ITR project and its research results to the public, and will ensure this working relationship is supported actively throughout the project. Accordingly, the Awardee agrees to inform the NSF Program Officer as far in advance as possible of any public announcements (including major journal articles) or press releases resulting from this project.

10. The Awardee agrees to include as part of the required Final Project Report a Summary of Major Project Achievements and a discussion of Lessons Learned.